A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information

Deepanshu Vasal and Achilleas Anastasopoulos

Abstract

We consider a finite horizon dynamic game with N selfish players who observe their types privately and take actions, which are publicly observed. Players' types evolve as conditionally independent Markov processes, conditioned on their current actions. Their actions and types jointly determine their instantaneous rewards. Since each player has a different information set, this is a dynamic game with asymmetric information and there is no known methodology to find perfect Bayesian equilibria (PBE) for such games in general. In this paper, we develop a methodology to obtain a class of PBE using a belief state based on players' common information. We first show that any expected reward profile that can be achieved by any general strategy profile can also be achieved by a policy based on players' private information and this belief state. With this structural result as our motivation, we develop our main result that provides a two-step backward-forward recursive algorithm to find a class of PBE of this game that are based on this belief state. We refer to such equilibria as *structured Bayesian perfect equilibria* (SPBE). The backward recursive part of this algorithm defines an equilibrium generating function. Each period in the backward recursion involves solving a fixed point equation on the space of probability simplexes for every possible belief on types. Using this function, equilibrium strategies and beliefs are generated through a forward recursion.

I. Introduction

There are many practical scenarios where strategic players with different sets of observations are involved in a time-evolving dynamical process such that their actions influence each others' payoffs. Such scenarios include repeated online advertisement auctions, wireless resource sharing, competing sellers and energy markets. In the case of repeated online advertisement auctions, advertisers place bids for locations on a website to sell a product. These bids are based on the value of that product, which is privately observed by an advertiser and past actions of everybody else, which are observed publicaly. Each advertiser's goal is to maximize its reward, which depends on the value of the products and on the actions taken by everybody else. A similar scenario can be considered for wireless resource sharing where players are allocated channels that interfere with each other. Each player privately observes its channel gain and takes actions, which may be the choice of modulation and coding scheme and also the transmission power. The reward here is the rate each player gets at time t, which is a function of everyone's channel gain and actions. Consider another scenario where different sellers compete to sell different but related goods which are complementary, substitutable or in general, with externalities. The true value of the goods is private information of a seller who, at each stage, takes an action to stock some amount of goods for sale. Her profit is based on some market mechanism (say through Walrasian prices) based on the true value of all the goods and their availability in the market, which depends on the actions of the other sellers. Each seller wants to maximize her own profit. Finally, a similar scenario also exists for energy markets where different suppliers (to their different end consumers) bid their estimated power outputs to an independent system operator (ISO) that forms the market mechanism to determine the prices assessed to the different suppliers. Each supplier wants to maximize its returns, which depend on its cost of production of energy, which is their private information, and the market-determined prices which depend on all the bids.

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 48105 USA e-mail: {dvasal, anastas} at umich.edu.

This work is supported in part by NSF grant CIF-1111061.

This paper has originally appeared on arXiv.org on August 26, 2015 as working paper 1508.06269.

Such dynamical systems with strategic players are modeled as dynamic games. In dynamic games with perfect and symmetric information, subgame perfect equilibrium (SPE) is an appropriate equilibrium concept [1], [2], [3] and there is a backward recursive algorithm to find all subgame perfect equilibria of such games. Maskin and Tirole in [4] introduced the concept of Markov perfect equilibrium (MPE) for dynamic games with perfect and symmetric information where equilibrium strategies are dependent on some payoff relevant state of the system rather than on the entire history. However, for games with asymmetric information, since players have different information sets in each period, they need to form a belief on the information sets of other players, based upon which they predict their strategies. As a result, SPE or MPE are not appropriate equilibrium concepts for such setting. There are several notions of equilibrium for such games, such as perfect Bayesian equilibrium (PBE), sequential equilibrium, trembling hand equilibrium [1], [3]. Each of these notions of equilibrium consists of a strategy and a belief profile of all players. The equilibrium strategies are optimal given the beliefs and the beliefs are derived from the equilibrium strategy profile and using Bayes' rule (whenever possible), with some equilibrium concepts requiring further refinements. Due to this circular argument of beliefs being consistent with strategies, which are in turn optimal given the beliefs, finding such equilibria is a difficult task. Moreover, strategies are function of histories, which belong to an ever-expanding space, and thus the space of optimization also becomes computationally intractable. There is no known methodology to find such equilibria for general dynamic games with asymmetric information.

In this paper, we consider a model where players observe their types privately and publicly observe the actions taken by other players at the end of each period. Their instantaneous rewards depend on everyones' types and actions. We provide a two-step algorithm involving a backward recursion followed by a forward recursion to construct a class of PBE for the dynamic game in consideration, which we call *structured perfect Bayesian equilibria* (SPBE). In these equilibria, players' strategies are based on their type and a set of beliefs on each type which is common to all players and lie in a time-invariant space. These beliefs on players' types form independent controlled Markov processes that together summarize the common information history and are updated individually and sequentially, based on corresponding agents' actions and (partial) strategies. The algorithm works as follows. In a backward recursive way, for each stage, the algorithm finds an equilibrium strategy function for all possible beliefs on types of the players which involves solving a fixed point equation on the space of probability simplexes. Then, the equilibrium strategies and beliefs are obtained through forward recursion by operating on the function obtained in the backward step. The SBPEs that are developed in this paper are analogous to the MPEs for dynamic games with perfect information in the sense that players choose their actions based on beliefs that depend on common information and have Markovian dynamics, where actions of a players are now partial functions from their private information to their action sets.

Related literature on this topic include [5], [6] and [7]. Nayyar et al. in [5], [6] consider a model of dynamic games with asymmetric information. There is an underlying controlled Markov process where players jointly observe part of the process and also make some observations privately. It is shown in [5], [6] that the considered game with asymmetric information, under certain assumptions, can be transformed to another game with symmetric information. Once this is established, a backward recursive algorithm is provided to find MPE of the transformed game, which are equivalently Nash equilibria of the transformed symmetric information game. For this strong equivalence to hold, authors in [5], [6] make a critical assumption in their model: based on the common information, a player's posterior beliefs about the system state and about other players' information are independent of the strategies used by the players in the past. Our model is different from the model considered in [5], [6]. We assume that the underlying state of the system has independent components, each constituting the type of a player. However, we do not make any assumption regarding update of beliefs and allow the common information based belief state to depend on players' strategies.

Ouyang et al. in [7] consider a dynamic oligopoly game with N strategic sellers of different goods and M strategic buyers. Each seller privately observes the valuation of their good, which is assumed to have independent Markovian dynamics, thus resulting in a dynamic game of asymmetric information. In each period, sellers post prices for their goods and buyers make decisions regarding buying the goods. Then a public signal indicating buyers experience is revealed which depends on sellers' valuation of the goods. Authors in [7] consider a policy-dependent common information based belief state based on which they define the concept of common information based equilibria. They show that for any given update function of this belief state, which is consistent with strategies of the players, if all other players play actions based on this common belief and their private information, then player i faces a Markov decision process (MDP) with respect to its action with state as common belief and its type. For every prior distribution, this defines a fixed point equation on belief update functions and strategies of all players. They provide necessary and sufficient conditions for common information based strategy profile and belief update functions to constitute PBE of the game; however they do no provide a systematic way to find such equilibria. In addition, because of the special structure of the reward function, the problem admits a degenerate solution where agents' strategies do not depend on their private information and therefore no signaling takes place. This allows existence of myopic, type-independent equilibrium policies (although other equilibria may also exist).

The paper is organized as follows. In section II, we present our model. In section III we present structural results that serve as motivation for SPBE. In section IV we present the main result by providing a two-step backward-forward recursive algorithm to construct a strategy profile and a sequence of beliefs and show that it is a PBE of the dynamic game considered. As an illustration, we apply this algorithm on a discrete version of an example from [3] on repeated public good game in Section V. We conclude in section VI. All proofs are presented in Appendices.

A. *Notation*

We use uppercase letters for random variables and lowercase for their realizations. For any variable, subscripts represent time indices and superscripts represent player identities. We use notation $-i$ to represent all players other than player i i.e. $-i = \{1, 2, \ldots i-1, i+1, \ldots, N\}$. We use notation $A_{t:t'}$ to represent vector $(A_t, A_{t+1}, \ldots A_{t'})$ when $t' \geq t$ or an empty vector if $t' < t$. We use A_t^{-i} to mean $(A_t^1, A_t^2, \ldots, A_t^{i-1}, A_t^{i+1} \ldots, A_t^N)$. We remove superscripts or subscripts if we want to represent the whole vector, for example A_t represents $(A_t^1, \ldots, A_t^N)$. In a similar vein, for any collection of sets $(\mathcal{X}^i)_{i\in\mathcal{N}}$, we denote $\times_{i\in\mathcal{N}}\mathcal{X}^i$ by $\mathcal{X}$. We denote the indicator function of any set A by $I_A(\cdot)$. For any finite set $\mathcal{S}$, $\mathcal{P}(\mathcal{S})$ represents the space of probability measures on $\mathcal{S}$ and $|\mathcal{S}|$ represents its cardinality. We denote by P^g (or E^g) the probability measure generated by (or expectation with respect to) strategy profile g. We denote the set of real numbers by $\mathbb{R}$. For a probabilistic strategy profile of players $(\beta_t^i)_{i\in\mathcal{N}}$ where probability of action a_t^i conditioned on $a_{1:t-1}x_{1;t}^i$ is given by $\beta_t^i(a_t^i|a_{1:t-1}, x_{1:t}^i)$, we use the short hand notation $\beta_t^{-i}(a_t^{-i}|a_{1:t-1}, x_{1:t}^{-i})$ to represent $\prod_{j\neq i}\beta_t^j(a_t^j|a_{1:t-1}, x_{1:t}^j)$. All equalities and inequalities involving random variables are to be interpreted in the *a.s.* sense.

II. Model

We consider a discrete-time dynamical system with N strategic players in the set $\mathcal{N} \triangleq \{1, 2, \ldots N\}$, over a time horizon $\mathcal{T} \triangleq \{1, 2, \ldots T\}$ and with perfect recall. There is a dynamic state of the system $X_t \triangleq (X_t^1, X_t^2, \ldots X_t^N)$, where $X_t^i \in \mathcal{X}^i$ is the type of player i at time t which is perfectly observed and is its private information. Types of the players evolve as conditionally independent, controlled Markov processes such that

$$P(x_1) = \prod_{i=1}^{N} Q_1^i(x_1^i) \tag{1a}$$

$$P(x_t|x_{1:t-1}, a_{1:t-1}) = P(x_t|x_{t-1}, a_{t-1}) \tag{1b}$$

$$= \prod_{i=1}^{N} Q_t^i(x_t^i|x_{t-1}^i, a_{t-1}), \tag{1c}$$

where Q_t^i are known kernels. Player i at time t takes action $a_t^i \in \mathcal{A}^i$ on observing $a_{1:t-1}$, which is common information among players, and $x_{1:t}^i$ which it observes privately. The sets $\mathcal{A}^i, \mathcal{X}^i$ are assumed to be finite. Let $g^i = (g_t^i)_{t\in\mathcal{T}}$ be a probabilistic strategy of player i where $g_t^i : \mathcal{A}^{t-1} \times (\mathcal{X}^i)^t \to \mathcal{P}(\mathcal{A}^i)$ such that player i plays action A_t^i according to $A_t^i \sim g_t^i(\cdot|a_{1:t-1}, x_{1:t}^i)$. Let $g \triangleq (g^i)_{i\in\mathcal{N}}$ be a strategy profile of all players. At the end of interval t, player i receives an instantaneous reward $R^i(x_t, a_t)$. The objective of player i is to maximize its total expected reward

$$J^{i,g} \triangleq \mathbb{E}^g \left\{ \sum_{t=1}^{T} R^i(X_t, A_t) \right\}. \tag{2}$$

With all players being strategic, this problem is modeled as a dynamic game $\mathfrak{D}$ with imperfect and asymmetric information, and with simultaneous moves.

III. Motivation for structured equilibria

In this section we present structural results for the considered dynamical process that serve as a motivation for finding SPBE of the underlying game $\mathfrak{D}$. Specifically, we define a belief state based on common information history and show that any reward profile that can be obtained through a general strategy profile can also be obtained through strategies that depend on this belief state and player's current type which is its private information. These structural results are inspired by the analysis of decentralized team problems, which serve as guiding principles to design our equilibrium strategies. While these structural results provide intuition and the required notation, they are not directly used in the proofs for finding SPBEs, later, in Section IV.

At any time t, player i has information $(a_{1:t-1}, x_{1:t}^i)$ where $a_{1:t-1}$ is the common information among players, and $x_{1:t}^i$ is the private information of player i. Since $(a_{1:t-1}, x_{1:t}^i)$ increases with time, any strategy of the form $A_t^i \sim g_t^i(\cdot|a_{1:t-1}, x_{1:t}^i)$ becomes unwieldy. Thus it is desirable to have an information state in a time-invariant space that succinctly summarizes $(a_{1:t-1}, x_{1:t}^i)$ and that can be sequentially updated. We first show in Fact 1 that given common information $a_{1:t-1}$ and its current type x_t^i, player i can discard its type history $x_{1:t-1}^i$ and play a strategy of the form $s_t^i(a_t^i|a_{1:t-1}, x_t^i)$. Then in Fact 2, we show that $a_{1:t-1}$ can be summarized through a belief π_t, defined as follows. For any strategy profile g, belief π_t on X_t, $\pi_t \in \mathcal{P}(\mathcal{X})$, is defined as $\pi_t(x_t) \triangleq P^g(X_t = x_t|a_{1:t-1})\ \forall x_t \in \mathcal{X}$. We also define the marginals $\pi_t^i(x_t^i) \triangleq P^g(x_t^i = x_t^i|a_{1:t-1})\ \forall x_t^i \in \mathcal{X}^i$.

For player i, we use notation g to denote a general policy of type $A_t^i \sim g_t^i(\cdot|a_{1:t-1}, x_{1:t}^i)$, notation s where $s_t^i : (\mathcal{A})^{t-1} \times \mathcal{X}^i \to \mathcal{P}(\mathcal{A}^i)$ to denote a policy of the form $s_t^i(a_t^i|a_{1:t-1}, x_t^i)$ and notation m where $m_t^i : \mathcal{P}(\times_{i\in\mathcal{N}}\mathcal{X}^i) \times \mathcal{X}^i \to \mathcal{P}(\mathcal{A}^i)$ to denote a policy of the form $m_t^i(a_t^i|\pi_t, x_t^i)$. It should be noted that since π_t is a function of random variables $a_{1:t-1}$, m policy is a special type of s policy, which in turn, is a special type of g policy.

Using the agent-by-agent approach [8], we show in Fact 1 that any expected reward profile of the players that can be achieved by any general strategy profile g can also be achieved by a strategy profile s.

Fact 1: Given a fixed strategy g^{-i} of all players other than player i and for any strategy g^i of player i, there exists a strategy s^i of player i such that

$$P^{s^i g^{-i}}(x_t, a_t) = P^{g^i g^{-i}}(x_t, a_t) \qquad \forall t \in \mathcal{T}, x_t \in \mathcal{X}, a_t \in \mathcal{A}, \tag{3}$$

which implies $J^{i,s^i g^{-i}} = J^{i,g^i g^{-i}}$.

Proof: See Appendix A. ■

Since any s^i policy is also a g^i type policy, the above fact can be iterated over all players which implies that for any g policy profile there exists an s policy profile that achieves the same reward profile i.e. $(J^{i,s})_{i\in\mathcal{N}} = (J^{i,g})_{i\in\mathcal{N}}$.

Policies of types s still have increasing domain due to increasing common information, $a_{1:t-1}$. In order to summarize this information, we take an equivalent view of the system dynamics through a common agent, as taken by Nayyar et al. in [9]. The common agent approach is a general approach that has been used extensively for dynamic team problems [10]–[13]. Using this approach, the problem can be equivalently described as follows: player i at time t observes $a_{1:t-1}$ and takes action γ_t^i, where $\gamma_t^i : \mathcal{X}^i \to \mathcal{P}(\mathcal{A}^i)$ is a partial (stochastic) function from its private information x_t^i to a_t^i of the form $\gamma_t^i(a_t^i|x_t^i)$. These actions are generated through some policy $\psi^i = (\psi_t^i)_{t\in\mathcal{T}}$, $\psi_t^i : \mathcal{A}^{t-1} \to \{\mathcal{X}^i \to \mathcal{P}(\mathcal{A}^i)\}$, that operates on the common information $a_{1:t-1}$ so that $\gamma_t^i = \psi_t^i[a_{1:t-1}]$. Then any policy of the form $A_t^i \sim s_t^i(\cdot|a_{1:t-1}, x_t^i)$ is equivalent to $A_t^i \sim \psi_t^i[a_{1:t-1}](\cdot|x_t^i)$ [9].

We call a player i's policy through common agent to be of type ψ^i if its actions γ_t^i are taken as $\gamma_t^i = \psi_t^i[a_{1:t-1}]$. We call a player i's policy through common agent to be of type θ^i where $\theta_t^i : \mathcal{P}(\mathcal{X}) \to \{\mathcal{X}^i \to \mathcal{P}(\mathcal{A}^i)\}$, if its actions γ_t^i are taken as $\gamma_t^i = \theta_t^i[\pi_t]$. A policy of type θ^i is also a policy of type ψ^i. There is a one-to-one correspondence between policies of type s^i and of type ψ^i and between policies of type m^i and of type θ^i.

In the following fact, we show that the space of profiles of type s is outcome-equivalent to the space of profiles of type m.

Fact 2: For any given strategy profile s of all players, there exists a strategy profile m such that

$$P^m(x_t, a_t) = P^s(x_t, a_t) \qquad \forall t \in \mathcal{T}, x_t \in \mathcal{X}, a_t \in \mathcal{A}, \tag{4}$$

which implies $(J^{i,m})_{i\in\mathcal{N}} = (J^{i,s})_{i\in\mathcal{N}}$. Furthermore π_t can be factorized as $\pi_t(x_t) = \prod_{i=1}^N \pi_t^i(x_t^i)$ where each π_t^i can be updated through an update function

$$\pi_{t+1}^i = \bar{F}(\pi_t^i, \gamma_t^i, a_t), \tag{5}$$

where $\bar{F}$ is independent of s.

Proof: See Appendix B. ■

The above two facts show that any reward profile that can be generated through policy profile of type g can also be generated through policy profile of type m. It should be noted that the construction of s^i, as in (28d), depends only on g^i, while the construction of m^i depends on the whole policy profile g and not just on g^i, since construction of θ^i depends on ψ in (40). Thus any unilateral deviation of player i in g policy profile does not necessarily translate to unilateral deviation of player i in the corresponding m policy profile. Therefore g being an equilibrium of the game (in some appropriate notion) does not necessitate the corresponding m also being an equilibrium.

As shown in the previous facts, due to the independence of types and their evolution as independent controlled Markov processes, for any strategy of the players, joint beliefs on types can be factorized as product of their marginals i.e. $\pi_t(x_t) = \prod_{i=1}^N \pi_t^i(x_t^i)$. Since in this paper, we only deal with such joint beliefs, to accentuate this independence structure, we define $\underline{\pi}_t \in \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{X}^i)$ as vector of marginal beliefs where $\underline{\pi}_t := (\pi_t^i)_{i\in\mathcal{N}}$. In the rest of the paper, we will use $\underline{\pi}_t$ instead of π_t whenever appropriate, where of course π_t can be constructed from $\underline{\pi}_t$. Similarly, we define vector of belief updates as $F(\underline{\pi}, \gamma, a) := (\bar{F}(\pi^i, \gamma^i, a))_{i\in\mathcal{N}}$. We also change the notation of policies of type m as $m_t^i : \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{X}^i) \times \mathcal{X}^i \to \mathcal{P}(\mathcal{A}^i)$ and common agent's policies of type θ as $\theta_t^i : \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{X}^i) \to \{\mathcal{X}^i \to \mathcal{P}(\mathcal{A}^i)\}$.

We end this section by noting that finding general PBEs of type g of the game $\mathfrak{D}$ would be a desirable goal, but due to the space of strategies growing exponentially with time, that would be computationally intractable. However Fact 1 suggests that strategies of type m form a class that is rich in the sense that they achieve every possible reward profile. Since these strategies are functions of beliefs π_t that lie in a time-invariant space and are easily updatable, equilibria of this type are potential candidates for computation through backward recursion. In this paper our goal is to devise an algorithm to find structured equilibria of type m of the dynamic game $\mathfrak{D}$.

IV. Algorithm for SPBE computation

A. Preliminaries

Any history of this game at which players take action is of the form $h_t = (a_{1:t-1}, x_{1:t})$. Let $\mathcal{H}_t$ be the set of such histories of the game at time t when players take action, $\mathcal{H} \overset{\triangle}{=} \cup_{t=0}^T \mathcal{H}_t$ be the set of all possible such histories. At any time t player i observes $h_t^i = (a_{1:t-1}, x_{1:t}^i)$ and all players together have $h_t^c = a_{1:t-1}$ as common history. Let $\mathcal{H}_t^i$ be the set of observed histories of player i at time t and $\mathcal{H}_t^c$ be the set of common histories at time t. An appropriate concept of equilibrium for such games is PBE [3], which consists of a pair (β^*, μ^*) of strategy profile $\beta^* = (\beta_t^{*,i})_{t\in\mathcal{T}, i\in\mathcal{N}}$ where $\beta_t^{*,i} : \mathcal{H}_t^i \to \mathcal{P}(\mathcal{A}^i)$ and a belief profile $\mu^* = ({}^i\mu_t^*)_{t\in\mathcal{T}, i\in\mathcal{N}}$ where ${}^i\mu_t^* : \mathcal{H}_t^i \to \mathcal{P}(\mathcal{H}_t)$ that satisfy sequential rationality so that $\forall i \in \mathcal{N}, t \in \mathcal{T}, h_t^i \in \mathcal{H}_t^i, \beta^i$

$$\mathbb{E}^{\beta^{*,i}\beta^{*,-i},\,\mu^*[h_t^i]}\left\{\sum_{n=t}^{T} R^i(X_n, A_n)\big|h_t^i\right\} \geq \mathbb{E}^{\beta^{i}\beta^{*,-i},\,\mu^*[h_t^i]}\left\{\sum_{n=t}^{T} R^i(X_n, A_n)\big|h_t^i\right\}, \tag{6}$$

and the beliefs satisfy some consistency conditions as described in [3, p. 331]. In general, a belief for player i at time t, ${}^i\mu_t^*$ is defined on history $h_t = (a_{1:t-1}, x_{1:t})$ given its private history $h_t^i = (a_{1:t-1}, x_{1:t}^i)$. Here player i's private history $h_t^i = (a_{1:t-1}, x_{1:t}^i)$ consists of a public part $h_t^c = a_{1:t-1}$ and a private part $x_{1:t}^i$. At any time t, the relevant uncertainty player i has is about other players' type x_t^{-i}. In our setting, due to independence of types, player i's current type x_t^i does not provide any information about x_t^{-i} as will be shown later. For this reason we consider beliefs that are functions of each agent's history h_t^i only through the common history h_t^c. Hence, for each agent i, its belief for each history $h_t^c = a_{1:t-1}$ is derived from a common belief $\mu_t^*[a_{1:t-1}]$ which itself factorizes into a product of marginals $\prod_{j\in\mathcal{N}} \mu_t^{*,j}[a_{1:t-1}]$, as will be shown later. Thus we can sufficiently use the system of beliefs, $\mu^* = (\mu_t^*)_{t\in\mathcal{T}}$ with $\mu_t^* : \mathcal{H}_t^c \to \mathcal{P}(\mathcal{X})$, with the understanding that agent i's belief on x_t^{-i} is $\mu_t^{*,-i}[a_{1:t-1}](x_t^{-i}) = \prod_{j\neq i} \mu_t^{*,j}[a_{1:t-1}](x_t^j)$. Under the above structure, all consistency conditions that are required for PBEs [3, p. 331] are automatically satisfied.

Structural results from Section III provide us motivation to study equilibria of the form $(m_t^i(a_t^i|\underline{\pi}_t, x_t^i))_{i\in\mathcal{N}}$, which are equivalent to policy profiles of the form $(\theta_t^i[\underline{\pi}_t](a_t^i|x_t^i))_{i\in\mathcal{N}}$ and have the advantage of being defined on a time-invariant space.

B. Backward Recursion

In this section, we define an equilibrium generating function $\theta = (\theta_t^i)_{i\in\mathcal{N}, t\in\mathcal{T}}$, where $\theta_t^i : \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{X}^i) \to \{\mathcal{X}^i \to \mathcal{P}(\mathcal{A}^i)\}$ and a sequence of functions $(V_t^i)_{i\in\mathcal{N}, t\in\{1,2,\ldots T+1\}}$, where $V_t^i : \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{X}^i) \times \mathcal{X}^i \to \mathbb{R}$, in a backward recursive way, as follows.

1. Initialize $\forall \underline{\pi}_{T+1} \in \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{X}^i), x_{T+1}^i \in \mathcal{X}^i$,

$$V_{T+1}^i(\underline{\pi}_{T+1}, x_{T+1}^i) \triangleq 0. \tag{7}$$

2. For $t = T, T-1, \ldots 1$, $\forall \underline{\pi}_t \in \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{X}^i), \pi_t = \prod_{i\in\mathcal{N}} \pi_t^i$, let $\theta_t[\underline{\pi}_t]$ be generated as follows. Set $\tilde{\gamma}_t = \theta_t[\underline{\pi}_t]$, where $\tilde{\gamma}_t$ is the solution, if it exists[1], of the following equation, $\forall i \in \mathcal{N}, x_t^i \in \mathcal{X}^i$,

$$\tilde{\gamma}_t^i(\cdot|x_t^i) \in \arg\max_{\gamma_t^i(\cdot|x_t^i)} \mathbb{E}^{\gamma_t^i(\cdot|x_t^i)\tilde{\gamma}_t^{-i},\,\pi_t}\left\{R^i(X_t, A_t) + V_{t+1}^i(F(\underline{\pi}_t, \tilde{\gamma}_t, A_t), X_{t+1}^i)\big|x_t^i\right\}, \tag{8}$$

where expectation in (8) is with respect to random variables $(X_t^{-i}, A_t, X_{t+1}^i)$ through the measure $\pi_t^{-i}(x_t^{-i})\gamma_t^i(a_t^i|x_t^i)\tilde{\gamma}_t^{-i}(a_t^{-i}|x_t^{-i})Q_{t+1}^i(x_{t+1}^i|x_t^i, a_t)$ and F is defined in the proof of Fact 2 and in particular Claim 5.
Furthermore, set

$$V_t^i(\underline{\pi}_t, x_t^i) \triangleq \mathbb{E}^{\tilde{\gamma}_t^i(\cdot|x_t^i)\tilde{\gamma}_t^{-i},\,\pi_t}\left\{R^i(X_t, A_t) + V_{t+1}^i(F(\underline{\pi}_t, \tilde{\gamma}_t, A_t), X_{t+1}^i)\big|x_t^i\right\}. \tag{9}$$

It should be noted that in (8), $\tilde{\gamma}_t^i$ is not the outcome of the maximization operation as in a best response equation similar to that of a Bayesian Nash equilibrium. Rather (8) has characteristics of a fixed point equation. This is because the maximizer $\tilde{\gamma}_t^i$ appears in both, the left-hand-side and the right-hand-side of

[1] Similar to the existence results shown in [7], in the special case where agent i's instantaneous reward does not depend on its private type x_t^i, the fixed point equation always has a type-independent solution $\tilde{\gamma}_t^i(\cdot)$ since it degenerates to a best-response-like equation.

the equation. This distinct construction allows the maximization operation to be done with respect to the variable $\gamma^i_t(\cdot|x^i_t)$ for every x^i_t separately as opposed to be done with respect to the whole function $\gamma^i_t(\cdot|\cdot)$, and is pivotal in the construction.

To highlight the significance of structure of (8), we contrast it with two alternate incorrect constructions.

(a) Following the common information approach as in decentralized team problems [9], instead of (8), suppose γ^i_t were constructed as equilibrium on common agents' actions γ_t, i.e. for a fixed $\underline{\pi}_t, \pi_t = \prod_{i\in\mathcal{N}} \pi^i_t, \forall i \in \mathcal{N}$,

$$\tilde{\gamma}^i_t \in \arg\max_{\gamma^i_t} \mathbb{E}^{\gamma^i_t\tilde{\gamma}^{-i}_t,\pi_t}\left\{R^i(X_t,A_t) + V^i_{t+1}(F(\underline{\pi}_t,\gamma^i_t\tilde{\gamma}^{-i}_t,A_t),X^i_{t+1})\right\}. \tag{10}$$

It should be noted that in (10), the argument of the maximization operation, γ^i_t, appears both, in generation of action A^i_t and in the update of the belief π_t. Moreover, (10) is not conditioned on x^i_t, the private information of player i, similar to the case in the corresponding team problem. This is because the common agent who does not observe the private information of the player i, averages out that information. While this averaging of private information works for the team problem whose objective is to maximize the total expected reward, for the case with strategic players, it is incompatible with the sequential rationality condition in (6), which requires conditioning on the entire history $(a_{1:t-1}, x^i_t)$ and not just the common information $a_{1:t-1}$.

If the private information is also conditioned on, the construction still remains invalid, as discussed next.

(b) Instead of (8), suppose γ^i_t were constructed as best response of player i to other players actions $\tilde{\gamma}^{-i}_t$, similar to a standard Bayesian Nash equilibrium. For a fixed $\underline{\pi}_t, \pi_t = \prod_{i\in\mathcal{N}} \pi^i_t, \forall i \in \mathcal{N}, x^i_t \in \mathcal{X}^i$,

$$\tilde{\gamma}^i_t \in \arg\max_{\gamma^i_t} \mathbb{E}^{\gamma^i_t(\cdot|x^i)\tilde{\gamma}^{-i}_t,\pi_t}\left\{R^i(X_t,A_t) + V^i_{t+1}(F(\underline{\pi}_t,\gamma^i_t\tilde{\gamma}^{-i}_t,A_t),X^i_{t+1})\big|x^i_t\right\}. \tag{11}$$

Then $\tilde{\gamma}^i_t$ would be a function of $\tilde{\gamma}^{-i}_t$ and x^i_t through a best response relation $\tilde{\gamma}^i_t \in BR^i_{x^i_t}(\tilde{\gamma}^{-i}_t)$, where $BR^i_{x^i_t}$ is appropriately defined from (11). Consequently, every component of the solution of the fixed point equation $(\tilde{\gamma}^i_t \in BR^i_{x^i_t}(\tilde{\gamma}^{-i}_t))_{x^i_t\in\mathcal{X}^i, i\in\mathcal{N}}$, if it existed, would be a function of the whole type profile x_t, resulting in a mapping $\tilde{\gamma}^i_t = \theta^i_t[\underline{\pi}_t, x_t]$. Since player i only observes its own type x^i_t, it would not be able to implement the corresponding $\tilde{\gamma}^i_t$ and therefore the construction would be invalid.

C. Forward Recursion

As discussed above, a pair of strategy and belief profile (β^*, μ^*) is a PBE if it satisfies (6). Based on θ defined above in (7)–(9), we now construct a set of strategies β^* and beliefs μ^* for the game $\mathfrak{D}$ in a forward recursive way, as follows[2]. As before, we will use the notation $\underline{\mu}^*_t[a_{1:t-1}] := (\mu^{*,i}_t[a_{1:t-1}])_{i\in\mathcal{N}}$ where $\mu^*_t[a_{1:t-1}]$ can be constructed from $\underline{\mu}^*_t[a_{1:t-1}]$ as $\mu^*_t[a_{1:t-1}](x_t) = \prod_{i=1}^N \mu^{*,i}_t[a_{1:t-1}](x^i_t)\ \forall a_{1:t-1} \in \mathcal{H}^c_t$ where $\mu^{*,i}_t[a_{1:t-1}]$ is a belief on x^i_t.

1. Initialize at time $t = 1$,

$$\mu^*_1[\phi](x_1) := \prod_{i=1}^N Q^i_1(x^i_1). \tag{12}$$

2. For $t = 1, 2 \dots T, \forall i \in \mathcal{N}, a_{1:t} \in \mathcal{H}^c_{t+1}, x^i_{1:t} \in (\mathcal{X}^i)^t$

$$\beta^{*,i}_t(a^i_t|a_{1:t-1}, x^i_{1:t}) = \beta^{*,i}_t(a^i_t|a_{1:t-1}, x^i_t) := \theta^i_t[\underline{\mu}^*_t[a_{1:t-1}]](a^i_t|x^i_t) \tag{13}$$

[2] As discussed in starting of Section IV, beliefs at time t are functions of each agent's history h^i_t only through the common history h^c_t and are the same for all agents.

and

$$\mu^{*,i}_{t+1}[a_{1:t}] := \bar{F}(\mu^{*,i}_t[a_{1:t-1}], \theta^i_t[\underline{\mu}^*_t[a_{1:t-1}]], a_t) \tag{14}$$

where $\bar{F}$ is defined in the proof of Fact 2 and in particular Claim 5.

We now state our main result.

Theorem 1: A strategy and belief profile (β^*, μ^*), constructed through backward/forward recursion algorithm described in section IV is a PBE of the game, i.e. $\forall i \in \mathcal{N}, t \in \mathcal{T}, a_{1:t-1} \in \mathcal{H}^c_t, x^i_{1:t} \in (\mathcal{X}^i)^t, \beta^i$,

$$\mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T},\, \mu^*_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^i(X_n, A_n)\big|a_{1:t-1}, x^i_{1:t}\right\} \geq \mathbb{E}^{\beta^{i}_{t:T}\beta^{*,-i}_{t:T},\, \mu^*_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^i(X_n, A_n)\big|a_{1:t-1}, x^i_{1:t}\right\}. \tag{15}$$

Proof: See Appendix C ■

An intuitive explanation for why all players are able to use a common belief is the following. The sequence of beliefs defined above serve two purposes. First, for any player i, it puts a belief on x^{-i}_t to compute an expectation on the current and future rewards. Secondly, it predicts the actions of the other players since their strategies are functions of these beliefs. Since for any strategy profile, x^i_t is conditionally independent of x^{-i}_t given the common history $a_{1:t-1}$ and since other players do not observe x^i_t, knowledge of x^i_t does not affect this belief and thus in our definition, all players can use the same belief μ^* which is independent of their private information.

Independence of types is a crucial assumption in proving the above result, which manifests itself in Lemma 2 in Appendix D, used in the proof of Theorem 1. This is because, at equilibrium, player i's reward-to-go at time t, conditioned on its type x^i_t, depends on its strategy at time t, β^i_t, only through its action a^i_t and is independent of the corresponding partial function $\beta^i_t(\cdot|a_{1:t-1}, \cdot)$. In other words, given x^i_t and a^i_t, player i's reward-to-go is independent of β^i_t. We discuss this in more detail below.

At equilibrium, all players observe past actions $a_{1:t-1}$ and update their belief π_t, which is the same as $\mu^*_t[a_{1:t-1}]$, through the equilibrium strategy profile β^*. Now suppose at time t, player i decides to unilaterally deviate to $\hat{\beta}^i_t$ at time t for some history $a_{1:t-1}$ keeping the rest of its strategy the same. Then other players still update their beliefs $(\pi_t)_{t\in\{t+1,\ldots T\}}$ same as before and take their actions through equilibrium strategy $\beta^{*,-i}_t$ operated on π_t and x^{-i}_t, whereas player i forms a new belief $\hat{\pi}_{t+1}$ on x_t which depends on strategy profile $\beta^*_{1:t-1}\hat{\beta}_t, \beta^{*,-i}_t$. Thus at time t player i would need both the beliefs $\pi_{t+1}, \hat{\pi}_{t+1}$ to compute its expected future reward; π_{t+1} to predict other players' actions and $\hat{\pi}_{t+1}$ to form a true belief on x_t based on its information. As it turns out, due to independence of types, $\hat{\pi}_{t+1}$ does not provide additional information to player i to compute its future expected reward and thus it can be discarded. Intuitively, this is so because the belief on type j, π^j_{t+1} is a function of strategy and action of player j till time t (as shown in Claim 1 in the proof of Theorem 1 in Appendix C); thus $\pi^{-i}_{t+1} = \hat{\pi}^{-i}_{t+1}$. Now since player i already observes its type x^i_t, its belief $\hat{\pi}^i_t$ on x^i_t does not provide any additional information to player i, and thus π_t (which is the same as $\mu^*_t[a_{1:t-1}]$) sufficiently computes future expected reward for player i. Also π_{t+1} is updated from π_t, $\beta^*_t(\cdot|a_{1:t-1}, \cdot)$ and a_t, and is independent of $\hat{\beta}^i_t$ given a^i_t. This implies player i can use the equilibrium strategy β^*_t to update its future belief, as used in (8). Then by construction of θ and specifically due to (8), player i does not gain by unilaterally deviating at time t keeping the remainder of its strategy the same.

Finally, we note that in the two-step backward-forward algorithm described above, once the equilibrium generating function θ is defined through backward recursion, the SPBEs can be generated through forward recursion for any prior distribution Q on types X. Since, in comparison to the backward recursion, the forward recursive part of the algorithm is computationally insignificant, the algorithm computes SPBEs for different prior distributions at the same time.

In the next section, we discuss an example to illustrate the methodology described above for the construction of SPBEs.

V. Illustrative example: A two stage public goods game

We consider a discrete version of Example 8.3 from [3, ch.8], which is an instance of a repeated public good game. There are two players who play a two period game. In each period t, they simultaneously decide whether to contribute to the period t public good, which is a binary decision $a_t^i \in \{0,1\}$ for player $i = 1,2$. Before the start of period 2, both players know the actions taken by them in period 1. For both periods, each player gets reward 1 if at least one of them contributed and 0 if none does. Player i's cost of contributing is x^i which is its private information. Both players believe that x^is are drawn independently and identically with probability distribution Q with support $\{x^L, x^H\}$; $0 < x^L < 1 < x^H$, such that $P^Q(X^i = x^H) = q$ where $0 < q < 1$.

This example is similar to our model where $N = 2, T = 2$ and reward for player i in period t is

$$R^i(x, a_t) = \begin{cases} a_t^{-i} & \text{if } a_t^i = 0 \\ 1 - x^i & \text{if } a_t^i = 1. \end{cases} \tag{16}$$

We will use the backward recursive algorithm, defined in Section IV, to find an SPBE of this game. For period $t = 1,2$ and for $i = 1,2$, the partial functions γ_t^i can equivalently be defined through scalars p_t^{iL} and p_t^{iH} such that $\gamma_t^i(1|x^L) = p_t^{iL}$, $\gamma_t^i(0|x^L) = 1 - p_t^{iL}$ and $\gamma_t^i(1|x^H) = p_t^{iH}$, $\gamma_t^i(0|x^H) = 1 - p_t^{iH}$, where $p_t^{iL}, p_t^{iH} \in [0,1]$. Henceforth, we will use p_t^{iL} and p_t^{iH} interchangeably with the corresponding γ_t^i.

For $t = 2$ and for any fixed $\underline{\pi}_2 = (\pi_2^1, \pi_2^2)$, where $\pi_2^i = \pi_2^i(x^H) \in [0,1]$ represents a probability measure on the event $\{X^i = x^H\}$, player i's reward is

$$\mathbb{E}^{\gamma_2}\{R_2^i(X, A_2)|\pi_2, X^i = x^L\} = (1 - p_2^{iL})\left((1 - \pi_2^{-i})p_2^{-iL} + \pi_2^{-i}p_2^{-iH}\right) + p_2^{iL}(1 - x^L), \tag{17a}$$

$$\mathbb{E}^{\gamma_2}\{R_2^i(X, A_2)|\pi_2, X^i = x^H\} = (1 - p_2^{iH})\left((1 - \pi_2^{-i})p_2^{-iL} + \pi_2^{-i}p_2^{-iH}\right) + p_2^{iH}(1 - x^H). \tag{17b}$$

Let $\tilde{\gamma}_2 = \theta_2[\underline{\pi}_2]$ and equivalently $(\tilde{p}_2^{1L}, \tilde{p}_2^{2L}, \tilde{p}_2^{1H}, \tilde{p}_2^{2H}) = \theta_2[\pi_2]$ be defined through the following fixed point equation, which is equivalent to (8). For $i = 1,2$

$$\tilde{p}_2^{iL} \in \arg\max_{p_2^{iL}} \quad (1 - p_2^{iL})\left((1 - \pi_2^{-i})\tilde{p}_2^{-iL} + \pi_2^{-i}\tilde{p}_2^{-iH}\right) + p_2^{iL}(1 - x^L), \tag{18a}$$

$$\tilde{p}_2^{iH} \in \arg\max_{p_2^{iH}} \quad (1 - p_2^{iH})\left((1 - \pi_2^{-i})\tilde{p}_2^{-iL} + \pi_2^{-i}\tilde{p}_2^{-iH}\right) + p_2^{iH}(1 - x^H). \tag{18b}$$

Since $1 - x^H < 0$, $\tilde{p}_2^{iH} = 0$ achieves the maximum in (18b). Thus (18a)–(18b) can be reduced to, $\forall i \in \{1,2\}$

$$\tilde{p}_2^{iL} \in \arg\max_{p_2^{iL}} \quad (1 - p_2^{iL})(1 - \pi_2^{-i})\tilde{p}_2^{-iL} + p_2^{iL}(1 - x^L). \tag{19}$$

This implies,

$$\tilde{p}_2^{iL} = \begin{cases} 0 & \text{if} \quad x^L > 1 - (1 - \pi_2^{-i})\tilde{p}_2^{-iL}, \\ 1 & \text{if} \quad x^L < 1 - (1 - \pi_2^{-i})\tilde{p}_2^{-iL}, \\ \text{arbitrary} & \text{if} \quad x^L = 1 - (1 - \pi_2^{-i})\tilde{p}_2^{-iL}. \end{cases} \tag{20}$$

The fixed point equation (20) has the following solutions,

1) $(\tilde{p}_2^{1L}, \tilde{p}_2^{2L}, \tilde{p}_2^{1H}, \tilde{p}_2^{2H}) = (0,1,0,0)$ for $\pi_2^1 \in [0,1], \pi_2^2 \le x^L$
   - $V_2^1(\pi_2, x^L) = 1 - \pi_2^2$
   - $V_2^1(\pi_2, x^H) = 1 - \pi_2^2$
   - $V_2^2(\pi_2, x^L) = 1 - x^L$
   - $V_2^2(\pi_2, x^H) = 0.$
2) $(\tilde{p}_2^{1L}, \tilde{p}_2^{2L}, \tilde{p}_2^{1H}, \tilde{p}_2^{2H}) = (1,0,0,0)$ for $\pi_2^1 \le x^L, \pi_2^1 \in [0,1]$
   - $V_2^1(\pi_2, x^L) = 1 - x^L$
   - $V_2^1(\pi_2, x^H) = 0$
   - $V_2^2(\pi_2, x^L) = 1 - \pi_2^1$

- $V_2^2(\pi_2, x^H) = 1 - \pi_2^1$.

3) $(\tilde{p}_2^{1L}, \tilde{p}_2^{2L}, \tilde{p}_2^{1H}, \tilde{p}_2^{2H}) = (1, 1, 0, 0)$ for $\pi_2^1 \geq x^L, \pi_2^2 \geq x^L$
   - $V_2^1(\pi_2, x^L) = 1 - x^L$
   - $V_2^1(\pi_2, x^H) = 1 - \pi_2^2$
   - $V_2^2(\pi_2, x^L) = 1 - x^L$
   - $V_2^2(\pi_2, x^H) = 1 - \pi_2^1$.
4) $(\tilde{p}_2^{1L}, \tilde{p}_2^{2L}, \tilde{p}_2^{1H}, \tilde{p}_2^{2H}) = (1, \tilde{p}_2^{2L}, 0, 0)$ for $\pi_2^1 = x^L, \pi_2^2 \in [0, 1]$ where $\tilde{p}_2^{2L} \in \left[0, \max\left\{1, \frac{1-x^L}{1-\pi_2^2}\right\}\right]$
   - $V_2^1(\pi_2, x^L) = 1 - x^L$
   - $V_2^1(\pi_2, x^H) = 1 - \pi_2^2.\tilde{p}_2^{2L}$
   - $V_2^2(\pi_2, x^L) = 1 - x^L$
   - $V_2^2(\pi_2, x^H) = 1 - \pi_2^1$.
5) $(\tilde{p}_2^{1L}, \tilde{p}_2^{2L}, \tilde{p}_2^{1H}, \tilde{p}_2^{2H}) = (\tilde{p}_2^{1L}, 1, 0, 0)$ for $\pi_2^1 \in [0, 1], \pi_2^2 = x^L$ where $\tilde{p}_2^{1L} \in \left[0, \max\left\{1, \frac{1-x^L}{1-\pi_2^1}\right\}\right]$
   - $V_2^1(\pi_2, x^L) = 1 - x^L$
   - $V_2^1(\pi_2, x^H) = 1 - \pi_2^2$
   - $V_2^2(\pi_2, x^L) = 1 - x^L$
   - $V_2^2(\pi_2, x^H) = 1 - \pi_2^1.\tilde{p}_2^{1L}$.
6) $(\tilde{p}_2^{1L}, \tilde{p}_2^{2L}, \tilde{p}_2^{1H}, \tilde{p}_2^{2H}) = (\frac{1-x^L}{1-\pi_2^1}, \frac{1-x^L}{1-\pi_2^2}, 0, 0)$ for $\pi_2^1 \leq x^L, \pi_2^2 \leq x^L$
   - $V_2^1(\pi_2, x^L) = 1 - x^L$
   - $V_2^1(\pi_2, x^H) = 1 - x^L$
   - $V_2^2(\pi_2, x^L) = 1 - x^L$
   - $V_2^2(\pi_2, x^H) = 1 - x^L$.

Figure 1 shows these solutions in the space of (π_2^1, π_2^2).

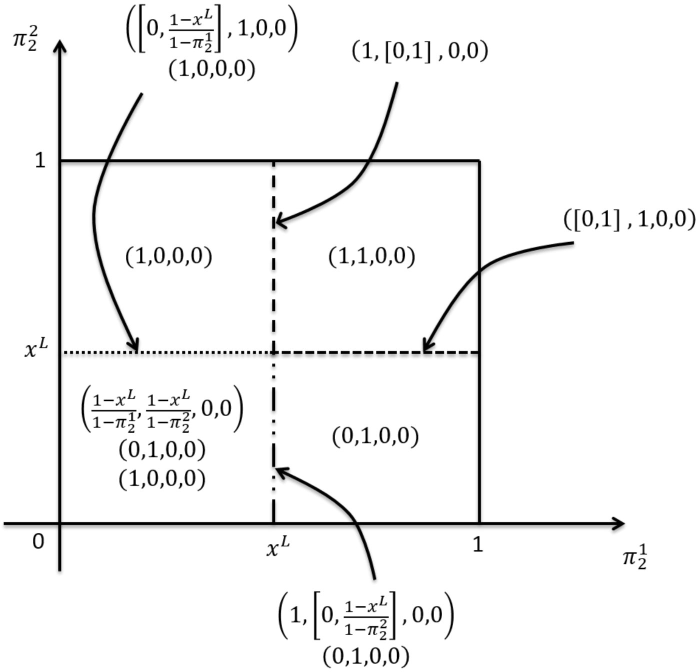
π_2^2
$\left(\left[0, \frac{1-x^L}{1-\pi_2^1}\right], 1,0,0\right)$
(1,0,0,0)
(1, [0,1] , 0,0)
1
([0,1] , 1,0,0)
(1,0,0,0)
(1,1,0,0)
x^L
$\left(\frac{1-x^L}{1-\pi_2^1}, \frac{1-x^L}{1-\pi_2^2}, 0,0\right)$
(0,1,0,0)
(1,0,0,0)
(0,1,0,0)
0
x^L
1
π_2^1
$\left(1, \left[0, \frac{1-x^L}{1-\pi_2^2}\right], 0,0\right)$
(0,1,0,0)

Fig. 1: Solutions of fixed point equation in (20)

Thus for any $\underline{\pi}_2$, there can exist multiple equilibria and correspondingly multiple $\theta_2[\underline{\pi}_2]$ can be defined.

For any particular θ_2, at $t=1$, the fixed point equation that needs to be solved is of the form, $\forall i \in \{1,2\}$

$$\begin{aligned}\tilde{p}_1^{iL} \in \arg\max_{p_1^{iL}} \quad & (1-p_1^{iL})\left((1-q)\tilde{p}_1^{-iL} + q\tilde{p}_1^{-iH} + \mathbb{E}^{\tilde{\gamma}_1}\{V_2^i(F(Q^2,\tilde{\gamma}_1,(0,A_1^{-i})),x^L)\}\right) \\ & + p_1^{iL}\left(1-x^L+\mathbb{E}^{\tilde{\gamma}_1}\{V_2^i(F(Q^2,\tilde{\gamma}_1,(1,A_1^{-i})),x^L)\}\right). \qquad (21a)\\ \tilde{p}_1^{iH} \in \arg\max_{p_1^{iH}} \quad & (1-p_1^{iH})\left((1-q)\tilde{p}_1^{-iL} + q\tilde{p}_1^{-iH} + \mathbb{E}^{\tilde{\gamma}_1}\{V_2^i(F(Q^2,\tilde{\gamma}_1,(0,A_1^{-i})),x^H)\}\right) \\ & + p_1^{iH}\left(1-x^H+\mathbb{E}^{\tilde{\gamma}_1}\{V_2^i(F(Q^2,\tilde{\gamma}_1,(1,A_1^{-i})),x^H)\}\right). \qquad (21b)\end{aligned}$$

where $F(Q^2,\tilde{\gamma},(A^1,A^2)) = \bar{F}(Q,\tilde{\gamma}^1,A^1)\bar{F}(Q,\tilde{\gamma}^2,A^2)$ and

$$\bar{F}(Q,\tilde{\gamma}_1^i,0) = \frac{q(1-\tilde{p}_1^{iH})}{q(1-\tilde{p}_1^{iH})+(1-q)(1-\tilde{p}_1^{iL})}, \qquad (22a)$$

$$\bar{F}(Q,\tilde{\gamma}_1^i,1) = \frac{q\tilde{p}_1^{iH}}{q\tilde{p}_1^{iH}+(1-q)\tilde{p}_1^{iL}}, \qquad (22b)$$

if the denominators in (22a)–(22b) are strictly positive, else $\bar{F}(Q,\tilde{\gamma}_1^i,A^i) = Q$ as in the proof of Fact 2, and in particular Claim 5. A solution of the fixed point equation in (21a)-(21b) defines $\theta_1[Q^2]$.

Using one such θ defined as follows, we find an SPBE of the game for $q=0.1, x^L=0.2, x^H=1.2$. We use $\theta_2[\underline{\pi}_2]$ as one possible set of solutions of (20), shown in Figure 2 and described below,

$$\theta_2[\pi_2] = (\tilde{p}_2^{1L},\tilde{p}_2^{2L},\tilde{p}_2^{1H},\tilde{p}_2^{2H}) = \begin{cases} (\frac{1-x^L}{1-\pi_2^1},\frac{1-x^L}{1-\pi_2^2},0,0) & \pi_2^1\in[0,x^L), \pi_2^2\in[0,x^L) \\ (1,0,0,0) & \pi_2^1\in[0,x^L], \pi_2^2\in[x^L,1] \\ (0,1,0,0) & \pi_2^1\in[x^L,1], \pi_2^2\in[0,x^L] \\ (1,1,0,0) & \pi_2^1\in(x^L,1], \pi_2^2\in(x^L,1]. \end{cases} \qquad (23)$$

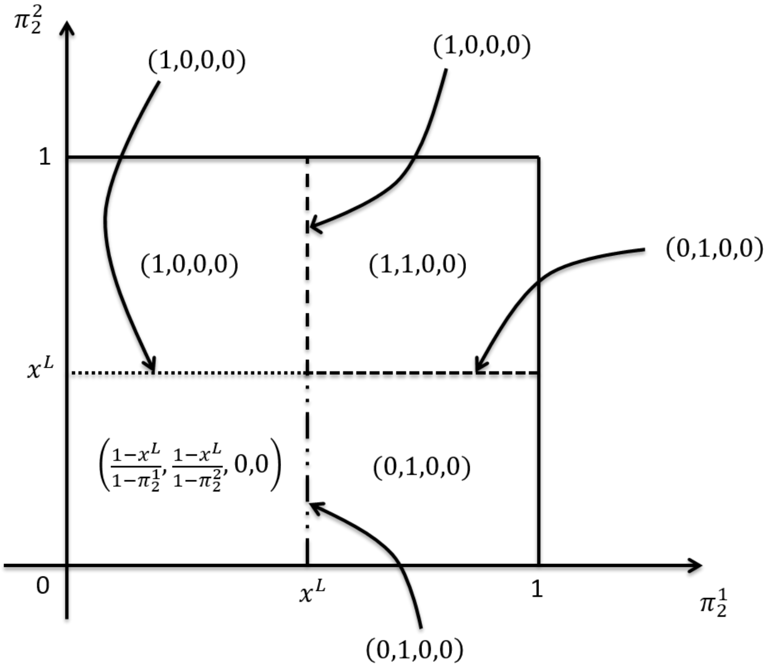

Fig. 2: $\theta_2[\underline{\pi}_2]$ described in (23)

Then, through iteration on the fixed point equation (21a)-(21b) and using the aforementioned $\theta_2[\underline{\pi}_2]$, we numerically find (and analytically verify) that $\theta_1[Q^2] = (\tilde{p}_1^{1L},\tilde{p}_1^{2L},\tilde{p}_1^{1H},\tilde{p}_1^{2H}) = (0,1,0,0)$ is a fixed point.

Thus

$$\begin{aligned}\beta_1^1(A_1^1 = 1|X^1 = x^L) = 0 \qquad \beta_1^2(A_1^2 = 1|X^2 = x^L) = 1\\ \beta_1^1(A_1^1 = 1|X^1 = x^H) = 0 \qquad \beta_1^2(A_1^2 = 1|X^2 = x^H) = 0\end{aligned}$$

with beliefs $\mu_2^*[00] = (q,1), \mu_2^*[01] = (q,0), \mu_2^*[10] = (q,1), \mu_2^*[11] = (q,0)$ and $(\beta_2^i(\cdot|a_1,\cdot))_{i\in\{1,2\}} = \theta_2[\mu_2^*[a_1]]$ is an SPBE of the game. In this equilibrium, player 2 at time $t=1$, contributes according to her type whereas player 1 never contributes, thus player 2 reveals her private information through her action whereas player 1 does not. Since θ_2 is symmetric, there also exists an (antisymmetric) equilibrium where at time $t=1$, players' strategies reverse i.e. player 2 never contributes and player 1 contributes according to her type. We also obtain a symmetric equilibrium where $\theta_1[Q^2] = (\frac{1-x^L}{(1-q)(1+x^L)}, \frac{1-x^L}{(1-q)(1+x^L)}, 0, 0)$ as a fixed point when $x^L > \frac{q}{2-q}$, resulting in beliefs $\mu_2^*[00] = (p,p), \mu_2^*[01] = (p,0), \mu_2^*[10] = (0,p), \mu_2^*[11] = (0,0)$ where $p = \frac{q(1+x^L)}{q(1+x^L)+(1-x^L)}$.

VI. Conclusion

In this paper, we study a class of dynamic games with asymmetric information where player i observes its true private type x_t^i and together with other players, observe past actions of everybody else. The types of the players evolve as conditionally independent, controlled Markov processes, conditioned on players current actions. We present a two-step backward-forward recursive algorithm to find SPBE of this game, where equilibrium strategies are function of a Markov belief state π_t, which depends on the common information, and current private types of the players. The backward recursive part of this algorithm defines an equilibrium generating function. Each period in backward recursion involves solving a fixed point equation on the space of probability simplexes for every possible belief on types. Then using this function, equilibrium strategies and beliefs are defined through a forward recursion.

In this paper we consider perfectly observable, independent dynamic types of the players. Future work includes considering types of players where players do not perfectly observe their types, rather they make noisy observations. In general, this methodology opens the door for finding PBEs for many applications, analytically or numerically, which was not feasible before. One such case would be dynamic LQG games where types evolve linearly with Gaussian noise and players incur quadratic cost.

Acknowledgment

The authors wish to acknowledge Vijay Subramanian for his contribution to the paper. Achilleas Anastasopoulos wishes to acknowledge Ashutosh Nayyar for the fruitful discussion and criticism of an early draft of this work presented during the ITA 2012 conference.

Appendix A
Proof of Fact 1

We prove this Fact in the following steps.

(a) In Claim 1, we prove that for any policy profile g and $\forall t \in \mathcal{T}$, $x_{1:t}^i$ for $i \in \mathcal{N}$ are conditionally independent given the common information $a_{1:t}$.

(b) In Claim 2, using Claim 1, we prove that for every fixed strategy g^{-i} of the players $-i$, $((a_{1:t-1}, x_t^i), a_t^i)_{t\in\mathcal{T}}$ is a controlled Markov process for player i.

(c) For a given policy g, we define a policy s^i of player i from g as $s_t^i(a_t^i|a_{1:t-1}, x_t^i) \overset{\triangle}{=} P^g(a_t^i|a_{1:t-1}, x_t^i)$.

(d) In Claim 3, we prove that the dynamics of this controlled Markov process $((x_t^i, a_{1:t-1}), a_t^i)_{t\in\mathcal{T}}$ under $(s^i g^{-i})$ are same as under g i.e. $P^{s^i g^{-i}}(x_t^i, x_{t+1}^i, a_{1:t}) = P^g(x_t^i, x_{t+1}^i, a_{1:t})$.

(e) In Claim 4, we prove that w.r.t. random variables (x_t, a_t), x_t^i is sufficient for player i's private information history $x_{1:t}^i$ i.e. $P^g(x_t, a_t|a_{1:t-1}, x_{1:t}^i, a_t^i) = P^{g^{-i}}(x_t, a_t|a_{1:t-1}, x_t^i, a_t^i)$.

(f) From (c), (d) and (e) we then prove the result of the Fact that $P^{s^i g^{-i}}(x_t, a_t) = P^g(x_t, a_t)$.

Claim 1: For any policy profile g and $\forall t$,

$$P^g(x_{1:t}|a_{1:t-1}) = \prod_{i=1}^{N} P^{g^i}(x^i_{1:t}|a_{1:t-1}) \tag{25}$$

Proof:

$$\begin{aligned}
P^g(x_{1:t}|a_{1:t-1}) &= \frac{P^g(x_{1:t}, a_{1:t-1})}{\sum_{\bar{x}_{1:t}} P^g(\bar{x}_{1:t}, a_{1:t-1})} &\text{(26a)}\\
&= \frac{\prod_{i=1}^N \left(Q^i_1(x^i_1) g^i_1(a^i_1|x^i_1) \prod_{n=2}^t Q^i_n(x^i_n|x^i_{n-1}, a_{n-1}) g^i_n(a^i_n|a_{1:n-1}, x^i_{1:n})\right)}{\sum_{\bar{x}_{1:t}} \prod_{i=1}^N \left(Q^i(\bar{x}^i_1) g^i_1(a^i_1|\bar{x}^i_1) \prod_{n=2}^t Q^i_n(\bar{x}^i_n|\bar{x}^i_{n-1}, a_{n-1}) g^i_n(a^i_n|a_{1:n-1}, \bar{x}^i_{1:n})\right)} &\text{(26b)}\\
&= \frac{\prod_{i=1}^N \left(Q^i_1(x^i_1) g^i_1(a^i_1|x^i_1) \prod_{n=2}^t Q^i_n(x^i_n|x^i_{n-1}, a_{n-1}) g^i_n(a^i_n|a_{1:n-1}, x^i_{1:n})\right)}{\prod_{i=1}^N \left(\sum_{\bar{x}^i_{1:t}} Q^i(\bar{x}^i_1) g^i_1(a^i_1|\bar{x}^i_1) \prod_{n=2}^t Q^i_n(\bar{x}^i_n|\bar{x}^i_{n-1}, a_{n-1}) g^i_n(a^i_n|a_{1:n-1}, \bar{x}^i_{1:n})\right)} &\text{(26c)}\\
&= \prod_{i=1}^N \frac{Q^i_1(x^i_1) g^i_1(a^i_1|x^i_1) \prod_{n=2}^t Q^i_n(x^i_n|x^i_{n-1}, a_{n-1}) g^i_n(a^i_n|a_{1:n-1}, x^i_{1:n})}{\sum_{\bar{x}^i_{1:t}} Q^i(\bar{x}^i_1) g^i_1(a^i_1|\bar{x}^i_1) \prod_{n=2}^t Q^i_n(\bar{x}^i_n|\bar{x}^i_{n-1}, a_{n-1}) g^i_n(a^i_n|a_{1:n-1}, \bar{x}^i_{1:n})} &\text{(26d)}\\
&= \prod_{i=1}^N P^{g^i}(x^i_{1:t}|a_{1:t-1}) &\text{(26e)}
\end{aligned}$$

■

Claim 2: For a fixed g^{-i}, $\{(a_{1:t-1}, x^i_t), a^i_t\}_t$ is a controlled Markov process with state $(a_{1:t-1}, x^i_t)$ and control action a^i_t.

Proof:

$$\begin{aligned}
&P^g(\tilde{a}_{1:t}, x^i_{t+1}|a_{1:t-1}, x^i_{1:t}, a^i_{1:t})\\
&= \sum_{x^{-i}_{1:t}} P^g(\tilde{a}_{1:t}, x^i_{t+1}, x^{-i}_{1:t}|a_{1:t-1}, x^i_{1:t}, a^i_t) &\text{(27a)}\\
&= \sum_{x^{-i}_{1:t}} P^g(\tilde{a}^{-i}_t, x^i_{t+1}, x^{-i}_{1:t}|a_{1:t-1}, x^i_{1:t}, a^i_t) I_{(a_{1:t-1}, a^i_t)}(\tilde{a}_{1:t-1}, \tilde{a}^i_t) &\text{(27b)}\\
&= \sum_{x^{-i}_{1:t}} P^{g^{-i}}(x^{-i}_{1:t}|a_{1:t-1}) \left(\prod_{j\neq i} g^j_t(\tilde{a}^j_t|a_{1:t-1}, x^j_{1:t})\right) Q^i_t(x^i_{t+1}|x^i_t, a^i_t, \tilde{a}^{-i}_t) I_{(a_{1:t-1}, a^i_t)}(\tilde{a}_{1:t-1}, \tilde{a}^i_t) &\text{(27c)}\\
&= P^{g^{-i}}(\tilde{a}_{1:t}, x^i_{t+1}|a_{1:t-1}, x^i_t, a^i_t), &\text{(27d)}
\end{aligned}$$

where (27c) follows from Claim 1 since $x^{-i}_{1:t}$ is conditionally independent of $x^i_{1:t}$ given $a_{1:t-1}$ and the corresponding probability is only a function of g^{-i}. ■

For any given policy profile g, we construct a policy s^i in the following way,

$$\begin{aligned}
s^i_t(a^i_t|a_{1:t-1}, x^i_t) &\triangleq P^g(a^i_t|a_{1:t-1}, x^i_t) &\text{(28a)}\\
&= \frac{\sum_{x^i_{1:t-1}} P^g(a^i_t, x^i_{1:t}|a_{1:t-1})}{\sum_{\tilde{a}^i_t} \sum_{\tilde{x}^i_{1:t-1}} P^g(\tilde{a}^i_t, \tilde{x}^i_{1:t-1} x^i_t|a_{1:t-1})} &\text{(28b)}\\
&= \frac{\sum_{x^i_{1:t-1}} P^{g^i}(x^i_{1:t}|a_{1:t-1}) g^i_t(a^i_t|a_{1:t-1}, x^i_{1:t})}{\sum_{\tilde{a}^i_t} \sum_{\tilde{x}^i_{1:t-1}} P^{g^i}(\tilde{x}^i_{1:t-1} x^i_t|a_{1:t-1}) g^i_t(\tilde{a}^i_t|a_{1:t-1}, \tilde{x}^i_{1:t-1} x^i_t)} &\text{(28c)}\\
&= P^{g^i}(a^i_t|a_{1:t-1}, x^i_t), &\text{(28d)}
\end{aligned}$$

where dependence of (28c) on only g^i is due to Claim 1.

Claim 3: The dynamics of the Markov process $\{(x_t^i, a_{1:t-1}), a_t^i\}_t$ under $(s^i g^{-i})$ are the same as under g i.e.

$$P^{s^i g^{-i}}(x_t^i, x_{t+1}^i, a_{1:t}) = P^g(x_t^i, x_{t+1}^i, a_{1:t}) \quad \forall t \tag{29}$$

Proof: We prove this by induction. Clearly,

$$P^g(x_1^i) = P^{s^i g^{-i}}(x_1^i) = Q_1^i(x_1^i) \tag{30}$$

Now suppose (29) is true for $t-1$ which also implies that the marginals $P^g(x_t^i, a_{1:t-1}) = P^{s^i g^{-i}}(x_t^i, a_{1:t-1})$. Then

$$\begin{align} P^g(x_t^i, a_{1:t-1}, x_{t+1}^i, a_t) &= P^g(x_t^i, a_{1:t-1}) P^g(a_t^i | a_{1:t-1}, x_t^i) P^g(x_{t+1}^i, a_{1:t} | x_t^i, a_{1:t-1}, a_t^i) \tag{31a}\\ &= P^{s^i g^{-i}}(x_t^i, a_{1:t-1}) s_t^i(a_t^i | a_{1:t-1}, x_t^i) P^{g^{-i}}(x_{t+1}^i, a_{1:t} | x_t^i, a_{1:t-1}, a_t^i) \tag{31b}\\ &= P^{s^i g^{-i}}(x_t^i, a_{1:t-1}, x_{t+1}^i, a_t) \tag{31c} \end{align}$$

where (31b) is true from induction hypothesis, definition of s^i in (28d) and since $\{(a_{1:t-1}, x_t^i), a_t^i\}_t$ is a controlled Markov process as proved in Claim 2 and its update kernel does not depend on policy g^i.This completes the induction step.

■

Claim 4: For any policy g,

$$P^g(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x_{1:t}^i, a_t^i) = P^{g^{-i}}(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x_t^i, a_t^i) \tag{32}$$

Proof:

$$P^g(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x_{1:t}^i, a_t^i) = I_{x_t^i, a_t^i}(\tilde{x}_t^i, \tilde{a}_t^i) P^g(\tilde{x}_t^{-i}, \tilde{a}_t^{-i} | a_{1:t-1}, x_{1:t}^i) \tag{33}$$

Now

$$\begin{align} P^g(\tilde{x}_t^{-i}, \tilde{a}_t^{-i} | a_{1:t-1}, x_{1:t}^i) &= \sum_{\tilde{x}_{1:t-1}^{-i}} P^g(\tilde{x}_{1:t}^{-i}, \tilde{a}_t^{-i} | a_{1:t-1}, x_{1:t}^i) \tag{34a}\\ &= \sum_{\tilde{x}_{1:t-1}^{-i}} P^g(\tilde{x}_{1:t}^{-i} | a_{1:t-1}, x_{1:t}^i) \left(\prod_{j \neq i} g_t^j(\tilde{a}_t^j | a_{1:t-1}, \tilde{x}_{1:t}^j) \right) \tag{34b}\\ &= \sum_{\tilde{x}_{1:t}^{-i}} P^{g^{-i}}(\tilde{x}_{1:t}^{-i} | a_{1:t-1}) \left(\prod_{j \neq i} g_t^j(\tilde{a}_t^j | a_{1:t-1}, \tilde{x}_{1:t}^j) \right) \tag{34c}\\ &= P^{g^{-i}}(\tilde{x}_t^{-i}, \tilde{a}_t^{-i} | a_{1:t-1}) \tag{34d} \end{align}$$

where (34c) follows from Claim 1.

Hence

$$\begin{align} P^g(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x_{1:t}^i, a_t^i) &= I_{x_t^i, a_t^i}(\tilde{x}_t^i, \tilde{a}_t^i) P^{g^{-i}}(\tilde{x}_t^{-i}, \tilde{a}_t^{-i} | a_{1:t-1}) \tag{35a}\\ &= P^{g^{-i}}(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x_t^i, a_t^i) \tag{35b} \end{align}$$

■

Finally,

$$P^g(\tilde{x}_t, \tilde{a}_t) = \sum_{a_{1:t-1}x^i_{1:t}a^i_t} P^g(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x^i_{1:t}, a^i_t) P^g(a_{1:t-1}, x^i_{1:t}, a^i_t) \tag{36a}$$

$$= \sum_{a_{1:t-1}x^i_{1:t},a^i_t} P^{g^{-i}}(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x^i_t, a^i_t) P^g(a_{1:t-1}, x^i_{1:t}, a^i_t) \tag{36b}$$

$$= \sum_{a_{1:t-1}x^i_t,a^i_t} P^{g^{-i}}(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x^i_t, a^i_t) P^g(a_{1:t-1}, x^i_t, a^i_t) \tag{36c}$$

$$= \sum_{a_{1:t-1}x^i_t,a^i_t} P^{g^{-i}}(\tilde{x}_t, \tilde{a}_t | a_{1:t-1}, x^i_t, a^i_t) P^{s^i g^{-i}}(a_{1:t-1}, x^i_t, a^i_t) \tag{36d}$$

$$= P^{s^i g^{-i}}(\tilde{x}_t, \tilde{a}_t). \tag{36e}$$

where (36b) follows from (32) in Claim 4 and (36d) from (29) in Claim 3.

Appendix B
(Proof of Fact 2)

For this proof we will assume the common agents strategies to be probabilistic as opposed to being deterministic, as was the case in section III. This means actions of the common agent, γ^i_t's are generated probabilistically from ψ^i as $\Gamma^i_t \cdot \psi^i_t(\cdot|a_{1:t-1})$, as opposed to being deterministically generated as $\gamma^i_t = \psi^i_t[a_{1:t-1}]$, as before. These two are equivalent ways of generating actions a^i_t from $a_{1:t-1}$ and x^i_t. We avoid using the probabilistic strategies of common agent throughout the main text for ease of exposition and because it conceptually does not affect the results.

Proof:

We prove this Fact in the following steps. We view this problem from the perspective of a common agent. Let ψ be the coordinator's policy corresponding to policy profile g. Let $\pi^i_t(x^i_t) = P^{\psi^i}(x^i_t|a_{1:t-1})$.

(a) In Claim 5, we show that π_t can be factorized as $\pi_t(x_t) = \prod_{i=1}^N \pi^i_t(x^i_t)$ where each π^i_t can be updated through an update function $\pi^i_{t+1} = \bar{F}(\pi^i_t, \gamma^i_t, a_t)$ and $\bar{F}$ is independent of common agent's policy ψ.

(b) In Claim 6, we prove that $(\Pi_t, \Gamma_t)_{t\in\mathcal{T}}$ is a controlled Markov process.

(c) We construct a policy profile θ from g such that $\theta_t(d\gamma_t|\pi_t) \overset{\triangle}{=} P^{\psi}(d\gamma_t|\pi_t)$.

(d) In Claim 7, we prove that dynamics of this Markov process $(\Pi_t, \Gamma_t)_{t\in\mathcal{T}}$ under θ is same as under ψ i.e. $P^{\theta}(d\pi_t, d\gamma_t, d\pi_{t+1}) = P^{\psi}(d\pi_t, d\gamma_t, d\pi_{t+1})$.

(e) In Claim 8, we prove that with respect to random variables (X_t, A_t), π_t can summarize common information $a_{1:t-1}$ i.e. $P^{\psi}(x_t, a_t|a_{1:t-1}, \gamma_t) = P(x_t, a_t|\pi_t, \gamma_t)$.

(f) From (c), (d) and (e) we that prove the result of the Fact that $P^{\psi}(x_t, a_t) = P^{\theta}(x_t, a_t)$ which is equivalent to $P^g(x_t, a_t) = P^m(x_t, a_t)$, where m is the policy profile of players corresponding to θ .

Claim 5: π_t can be factorized as $\pi_t(x_t) = \prod_{i=1}^N \pi^i_t(x^i_t)$ where each π^i_t can be updated through an update function $\pi^i_{t+1} = \bar{F}(\pi^i_t, \gamma^i_t, a_t)$ and $\bar{F}$ is independent of common agent's policy ψ. We also say $\underline{\pi}_{t+1} = F(\underline{\pi}_t, \gamma_t, a_t)$.

Proof:

We prove this by induction. Since $\pi_1(x_1) = \prod_{i=1}^N Q^i_t(x^i_1)$, the base case is verified. Now suppose $\pi_t =$

$\prod_{i=1}^{N} \pi_t^i$. Then,

$$\begin{aligned}
\pi_{t+1}(x_{t+1}) &= P^{\psi}(x_{t+1}|a_{1:t}, \gamma_{1:t+1}) && \text{(37a)}\\
&= P^{\psi}(x_{t+1}|a_{1:t}, \gamma_{1:t}) && \text{(37b)}\\
&= \frac{\sum_{x_t} P^{\psi}(x_t, a_t, x_{t+1}|a_{1:t-1}, \gamma_{1:t})}{\sum_{\tilde{x}_{t+1}\tilde{x}_t} P^{\psi}(\tilde{x}_t, \tilde{x}_{t+1}, a_t|a_{1:t-1}, \gamma_{1:t})} && \text{(37c)}\\
&= \frac{\sum_{x_t} \pi_t(x_t) \prod_{i=1}^{N} \gamma_t^i(a_t^i|x_t^i) Q_t^i(x_{t+1}^i|x_t^i, a_t)}{\sum_{\tilde{x}_t\tilde{x}_{t+1}} \pi_t(\tilde{x}_t) \prod_{i=1}^{N} \gamma_t^i(a_t^i|\tilde{x}_t^i) Q_t^i(\tilde{x}_{t+1}^i|\tilde{x}_t^i, a_t)} && \text{(37d)}\\
&= \prod_{i=1}^{N} \frac{\sum_{x_t^i} \pi_t^i(x_t^i)\gamma_t^i(a_t^i|x_t^i) Q_t^i(x_{t+1}^i|x_t^i, a_t)}{\sum_{\tilde{x}_t^i} \pi_t^i(\tilde{x}_t^i)\gamma_t^i(a_t^i|\tilde{x}_t^i)}, && \text{(37e)}\\
&= \prod_{i=1}^{N} \pi_{t+1}^i(x_{t+1}^i) && \text{(37f)}
\end{aligned}$$

where (37e) follows from induction hypothesis. It is assumed in (37c)-(37e) that the denominator is not 0. If denominator corresponding to any γ_t^i is zero, we define

$$\pi_{t+1}^i(x_{t+1}^i) = \sum_{x_t^i} \pi_t^i(x_t^i) Q_t^i(x_{t+1}^i|x_t^i, a_t), \tag{38}$$

where π_{t+1} still satisfies (37f). Thus $\pi_{t+1}^i = \bar{F}(\pi_t^i, \gamma_t^i, a_t)$ and $\underline{\pi}_{t+1} = F(\underline{\pi}_t, \gamma_t, a_1)$ where $\bar{F}$ and F are appropriately defined from above. ■

Claim 6: $(\Pi_t, \Gamma_t)_{t\in\mathcal{T}}$ is a controlled Markov process with state Π_t and control action Γ_t

Proof:

$$\begin{aligned}
P^{\psi}(d\pi_{t+1}|\pi_{1:t}, \gamma_{1:t}) &= \sum_{a_t, x_t} P^{\psi}(d\pi_{t+1}, a_t, x_t|\pi_{1:t}, \gamma_{1:t}) && \text{(39a)}\\
&= \sum_{a_t, x_t} P^{\psi}(x_t|\pi_{1:t}, \gamma_{1:t}) \left\{ \prod_{i=1}^{N} \gamma_t^i(a_t^i|x_t^i) \right\} I_{F(\pi_t, \gamma_t, a_t)}(\pi_{t+1}) && \text{(39b)}\\
&= \sum_{a_t, x_t} \pi_t(x_t) \left\{ \prod_{i=1}^{N} \gamma_t^i(a_t^i|x_t^i) \right\} I_{F(\pi_t, \gamma_t, a_t)}(\pi_{t+1}) && \text{(39c)}\\
&= P(d\pi_{t+1}|\pi_t, \gamma_t). && \text{(39d)}
\end{aligned}$$

■

For any given policy profile ψ, we construct policy profile θ in the following way.

$$\theta_t(d\gamma_t|\pi_t) \overset{\triangle}{=} P^{\psi}(d\gamma_t|\pi_t). \tag{40}$$

Claim 7:

$$P^{\psi}(d\pi_t, d\gamma_t, d\pi_{t+1}) = P^{\theta}(d\pi_t, d\gamma_t, d\pi_{t+1}) \quad \forall t \in \mathcal{T}. \tag{41}$$

Proof: We prove this by induction. For $t = 1$,

$$P^{\psi}(d\pi_1) = P^{\theta}(d\pi_1) = I_Q(\pi_1). \tag{42}$$

Now suppose $P^\psi(d\pi_t) = P^\theta(d\pi_t)$ is true for t, then

$$P^\psi(d\pi_t, d\gamma_t, d\pi_{t+1}) = P^\psi(d\pi_t)P^\psi(d\gamma_t|\pi_t)P^\psi(d\pi_{t+1}|\pi_t\gamma_t) \tag{43a}$$

$$= P^\theta(d\pi_t)\theta_t(d\gamma_t|\pi_t)P(d\pi_{t+1}|\pi_t, \gamma_t) \tag{43b}$$

$$= P^\theta(d\pi_t, d\gamma_t, d\pi_{t+1}). \tag{43c}$$

where (43b) is true from induction hypothesis, definition of θ in (40) and since $(\Pi_t, \Gamma_t)_{t\in\mathcal{T}}$ is a controlled Markov process as proved in Claim 6 and thus its update kernel does not depend on policy ψ. This completes the induction step. ■

Claim 8: For any policy ψ,

$$P^\psi(x_t, a_t|a_{1:t-1}, \gamma_t) = P(x_t, a_t|\pi_t, \gamma_t). \tag{44}$$

Proof:

$$P^\psi(x_t, a_t|a_{1:t-1}, \gamma_t) = P^\psi(x_t|a_{1:t-1}, \gamma_t)\prod_{i\in\mathcal{N}}\gamma^i_t(a^i_t|x^i_t) \tag{45a}$$

$$= \pi_t(x_t)\prod_{i\in\mathcal{N}}\gamma^i_t(a^i_t|x^i_t) \tag{45b}$$

$$= P(x_t, a_t|\pi_t, \gamma_t). \tag{45c}$$

■

Finally,

$$P^\psi(x_t, a_t) = \sum_{a_{1:t-1},\gamma_t} P^\psi(x_t, a_t|a_{1:t-1}, \gamma_t)P^\psi(a_{1:t-1}, \gamma_t) \tag{46a}$$

$$= \sum_{a_{1:t-1}\gamma_t} P(x_t, a_t|\pi_t, \gamma_t)P^\psi(a_{1:t-1}, \gamma_t) \tag{46b}$$

$$= \sum_{\pi_t,\gamma_t} P(x_t, a_t|\pi_t, \gamma_t)P^\psi(\pi_t, \gamma_t) \tag{46c}$$

$$= \sum_{\pi_t,\gamma_t} P(x_t, a_t|\pi_t, \gamma_t)P^\theta(\pi_t, \gamma_t) \tag{46d}$$

$$= P^\theta(x_t, a_t). \tag{46e}$$

where (46b) follows from (44), (46c) is change of variable and (46d) from (41).

■

APPENDIX C
(PROOF OF THEOREM 1)

Proof: We prove (15) using induction and from results in Lemma 1, 2 and 3 proved in Appendix D. For base case at $t = T$, $\forall i \in \mathcal{N}, (a_{1:T-1}, x^i_{1:T}) \in \mathcal{H}^i_T, \beta^i$

$$\mathbb{E}^{\beta_T^{*,i}\beta_T^{*,-i},\mu_T^*[a_{1:T-1}]}\left\{R^i(X_T, A_T)\big|a_{1:T-1}, x^i_{1:T}\right\} = V^i_T(\underline{\mu}^*_T[a_{1:T-1}], x^i_T) \tag{47a}$$

$$\geq \mathbb{E}^{\beta^i_T\beta_T^{*,-i},\mu_T^*[a_{1:T-1}]}\left\{R^i(X_T, A_T)\big|a_{1:T-1}, x^i_{1:T}\right\}. \tag{47b}$$

where (47a) follows from Lemma 3 and (47b) follows from Lemma 1 in Appendix D.

Let the induction hypothesis be that for $t+1$, $\forall i \in \mathcal{N}, a_{1:t} \in \mathcal{H}^c_{t+1}, x^i_{1:t+1} \in (\mathcal{X}^i)^{t+1}, \beta^i$,

$$\mathbb{E}^{\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T},\,\mu^*_{t+1}[a_{1:t}]}\left\{\sum_{n=t+1}^{T} R^i(X_n,A_n)\big|a_{1:t},x^i_{1:t+1}\right\} \tag{48a}$$

$$\geq \mathbb{E}^{\beta^{i}_{t+1:T}\beta^{*,-i}_{t+1:T},\,\mu^*_{t+1}[a_{1:t}]}\left\{\sum_{n=t+1}^{T} R^i(X_n,A_n)\big|a_{1:t},x^i_{1:t+1}\right\}. \tag{48b}$$

Then $\forall i \in \mathcal{N}, (a_{1:t-1}, x^i_{1:t}) \in \mathcal{H}^i_t, \beta^i$, we have

$$\mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T},\,\mu^*_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^i(X_n,A_n)\big|a_{1:t-1},x^i_{1:t}\right\}$$

$$= V^i_t(\underline{\mu}^*_t[a_{1:t-1}], x^i_t) \tag{49a}$$

$$\geq \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\,\mu^*_t[a_{1:t-1}]}\left\{R^i(X_t,A_t) + V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t-1}A_t], X^i_{t+1})\big|a_{1:t-1},x^i_{1:t}\right\} \tag{49b}$$

$$= \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\,\mu^*_t[a_{1:t-1}]}\Big\{R^i(X_t,A_t)+ \mathbb{E}^{\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T},\,\mu^*_{t+1}[a_{1:t-1},A_t]}\left\{\sum_{n=t+1}^{T} R^i(X_n,A_n)\big|a_{1:t-1},A_t,x^i_{1:t+1}\right\}\big|a_{1:t-1},x^i_{1:t}\Big\} \tag{49c}$$

$$\geq \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\,\mu^*_t[a_{1:t-1}]}\Big\{R^i(X_t,A_t)+ \mathbb{E}^{\beta^{i}_{t+1:T}\beta^{*,-i}_{t+1:T}\mu^*_{t+1}[a_{1:t-1},A_t]}\left\{\sum_{n=t+1}^{T} R^i(X_n,A_n)\big|a_{1:t-1},A_t,x^i_{1:t},X^i_{t+1}\right\}\big|a_{1:t-1},x^i_{1:t}\Big\} \tag{49d}$$

$$= \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\,\mu^*_t[a_{1:t-1}]}\left\{R^i(X_t,A_t)+ \mathbb{E}^{\beta^{i}_{t:T}\beta^{*,-i}_{t:T}\mu^*_{t}[a_{1:t-1}]}\left\{\sum_{n=t+1}^{T} R^i(X_n,A_n)\big|a_{1:t-1},A_t,x^i_{1:t},X^i_{t+1}\right\}\big|a_{1:t-1},x^i_{1:t}\right\} \tag{49e}$$

$$= \mathbb{E}^{\beta^{i}_{t:T}\beta^{*,-i}_{t:T}\,\mu^*_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^i(X_n,A_n)\big|a_{1:t-1},x^i_{1:t}\right\}, \tag{49f}$$

where (49a) follows from Lemma 3, (49b) follows from Lemma 1, (49c) follows from Lemma 3, (49d) follows from induction hypothesis in (48b) and (49e) follows from Lemma 2. Moreover, construction of θ in (8), and consequently definition of β^* in (13) are pivotal for (49e) to follow from (49d).

We note that μ^* satisfies the consistency condition of [3, p. 331] from the fact that (a) for all t and for every common history $a_{1:t-1}$, all players use the same belief $\mu^*_t[a_{1:t-1}]$ on x_t and (b) the belief μ^*_t can be factorized as $\mu^*_t[a_{1:t-1}] = \prod_{i=1}^N \mu^{*,i}_t[a_{1:t-1}]$ $\forall a_{1:t-1} \in \mathcal{H}^c_t$ where $\mu^{*,i}_t$ is updated through Bayes' rule ($\bar{F}$) as in Claim 5 in Appendix B.

■

Appendix D

Lemma 1: $\forall t \in \mathcal{T}, i \in \mathcal{N}, (a_{1:t-1}, x^i_{1:t}) \in \mathcal{H}^i_t, \beta^i_t$

$$V^i_t(\underline{\mu}^*_t[a_{1:t-1}], x^i_t) \geq \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\,\mu^*_t[a_{1:t-1}]}\left\{R^i(X_t,A_t) + V^i_{t+1}(F(\underline{\mu}^*_t[a_{1:t-1}], \beta^*_t(\cdot|a_{1:t-1},\cdot), A_t), X^i_{t+1})\big|a_{1:t-1},x^i_{1:t}\right\}. \tag{50}$$

Proof: We prove this Lemma by contradiction.

Suppose the claim is not true for t. This implies $\exists i, \hat{\beta}^i_t, \hat{a}_{1:t-1}, \hat{x}^i_{1:t}$ such that

$$\mathbb{E}^{\hat{\beta}^i_t\beta^{*,-i}_t,\,\mu^*_t[\hat{a}_{1:t-1}]}\left\{R^i(X_t,A_t) + V^i_{t+1}(F(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1},\cdot), A_t), X^i_{t+1})\big|\hat{a}_{1:t-1},\hat{x}^i_{1:t}\right\} > V^i_t(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \hat{x}^i_t). \tag{51}$$

We will show that this leads to a contradiction.

Construct $\hat{\gamma}_t^i(a_t^i|x_t^i) = \begin{cases} \hat{\beta}_t^i(a_t^i|\hat{a}_{1:t-1}, \hat{x}_{1:t}^i) & x_t^i = \hat{x}_t^i \\ \text{arbitrary} & \text{otherwise.} \end{cases}$

Then for $\hat{a}_{1:t-1}, \hat{x}_{1:t}^i$, we have

$$V_t^i(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \hat{x}_t^i) \tag{52a}$$

$$= \max_{\gamma_t^i(\cdot|\hat{x}_t^i)} \mathbb{E}^{\gamma_t^i(\cdot|\hat{x}_t^i)\beta_t^{*,-i},\, \mu_t^*[\hat{a}_{1:t-1}]} \left\{ R^i(\hat{x}_t^i x_t^{-i}, a_t) + V_{t+1}^i(F(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \beta_t^*(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), X_{t+1}^i) \big| \hat{x}_t^i \right\}, \tag{52b}$$

$$\geq \mathbb{E}^{\hat{\gamma}_t^i(\cdot|\hat{x}_t^i)\beta_t^{*,-i},\, \mu_t^*[\hat{a}_{1:t-1}]} \left\{ R^i(X_t, A_t) + V_{t+1}^i(F(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \beta_t^*(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), X_{t+1}^i) \big| \hat{x}_t^i \right\}$$

$$= \sum_{x_t^{-i}, a_t, x_{t+1}} \left\{ R^i(\hat{x}_t^i x_t^{-i}, a_t) + V_{t+1}^i(F(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \beta_t^*(\cdot|\hat{a}_{1:t-1}, \cdot), a_t), x_{t+1}^i) \right\} \times$$
$$\mu_t^{*,-i}[\hat{a}_{1:t-1}](x_t^{-i}) \hat{\gamma}_t^i(a_t^i|\hat{x}_t^i) \beta_t^{*,-i}(a_t^{-i}|\hat{a}_{1:t-1}, x_t^{-i}) Q_t^i(x_{t+1}^i|\hat{x}_t^i, a_t) \tag{52c}$$

$$= \sum_{x_t^{-i}, a_t, x_{t+1}} \left\{ R^i(\hat{x}_t^i x_t^{-i}, a_t) + V_{t+1}^i(F(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \beta_t^*(\cdot|\hat{a}_{1:t-1}, \cdot), a_t), x_{t+1}^i) \right\} \times$$
$$\mu_t^{*,-i}[\hat{a}_{1:t-1}](x_t^{-i}) \hat{\beta}_t^i(a_t^i|\hat{a}_{1:t-1}, \hat{x}_{1:t}^i) \beta_t^{*,-i}(a_t^{-i}|\hat{a}_{1:t-1}, x_t^{-i}) Q_t^i(x_{t+1}^i|\hat{x}_t^i, a_t) \tag{52d}$$

$$= \mathbb{E}^{\hat{\beta}_t^i \beta_t^{*,-i},\, \mu_t^*[\hat{a}_{1:t-1}]} \left\{ R^i(\hat{x}_t^i x_t^{-i}, a_t) + V_{t+1}^i(F(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \beta_t^*(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), X_{t+1}^i) \big| \hat{a}_{1:t-1}, \hat{x}_{1:t}^i \right\} \tag{52e}$$

$$> V_t^i(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \hat{x}_t^i) \tag{52f}$$

where (52b) follows from definition of V_t^i in (9), (52d) follows from definition of $\hat{\gamma}_t^i$ and (52f) follows from (51). However this leads to a contradiction. ■

Lemma 2: $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t}, x_{1:t+1}^i) \in \mathcal{H}_{t+1}^i$ and β_t^i

$$\mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*,-i},\, \mu_t^*[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \big| a_{1:t}, x_{1:t+1}^i \right\} = \mathbb{E}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*,-i},\, \mu_{t+1}^*[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \big| a_{1:t}, x_{1:t+1}^i \right\}. \tag{53}$$

Thus the above quantities do not depend on β_t^i.

Proof: Essentially this claim stands on the fact that $\mu_{t+1}^{*,-i}[a_{1:t}]$ can be updated from $\mu_t^{*,-i}[a_{1:t-1}], \beta_t^{*,-i}$ and a_t, as $\mu_{t+1}^{*,-i}[a_{1:t}] = \prod_{j \neq i} \bar{F}(\mu_t^{*,-i}[a_{1:t-1}], \beta_t^{*,-i}, a_t)$ as in Claim 5. Since the above expectations involve random variables $X_{t+1}^{-i}, A_{t+1:T}, X_{t+2:T}$, we consider $P^{\beta_{t:T}^i \beta_{t:T}^{*,-i},\, \mu_t^*[a_{1:t-1}]}(x_{t+1}^{-i}, a_{t+1:T}, x_{t+2:T} \big| a_{1:t}, x_{1:t+1}^i)$.

$$P^{\beta_{t:T}^i \beta_{t:T}^{*,-i},\, \mu_t^*[a_{1:t-1}]}(x_{t+1}^{-i}, a_{t+1:T}, x_{t+2:T} \big| a_{1:t}, x_{1:t+1}^i)$$
$$= \frac{\sum_{x_t^{-i}} P^{\beta_{t:T}^i \beta_{t:T}^{*,-i},\, \mu_t^*[a_{1:t-1}]}(x_t^{-i}, a_t, x_{t+1}, a_{t+1:T}, x_{t+2:T} \big| a_{1:t-1}, x_{1:t}^i)}{\sum_{\tilde{x}_t^{-i}} P^{\beta_{t:T}^i \beta_{t:T}^{*,-i},\, \mu_t^*[a_{1:t-1}]}(\tilde{x}_t^{-i}, a_t, x_{t+1}^i \big| a_{1:t-1}, x_{1:t}^i)} \tag{54a}$$

We consider the numerator and the denominator separately. The numerator in (54a) is given by

$$Nr = \sum_{x_t^{-i}} P^{\beta_{t:T}^i \beta_{t:T}^{*,-i},\, \mu_t^*[a_{1:t-1}]}(x_t^{-i} \big| a_{1:t-1}, x_{1:t}^i) \beta_t^i(a_t^i|a_{1:t-1}, x_{1:t}^i) \beta_t^{*,-i}(a_t^{-i}|a_{1:t-1}, x_t^{-i}) Q(x_{t+1}|x_t, a_t)$$
$$P^{\beta_{t:T}^i \beta_{t:T}^{*,-i},\, \mu_t^*[a_{1:t-1}]}(a_{t+1:T}, x_{t+2:T} \big| a_{1:t}, x_{1:t-1}^i, x_{t:t+1}) \tag{54b}$$

$$= \sum_{x_t^{-i}} \mu_t^{*,-i}[a_{1:t-1}](x_t^{-i}) \beta_t^i(a_t^i|a_{1:t-1}, x_{1:t}^i) \beta_t^{*,-i}(a_t^{-i}|a_{1:t-1}, x_t^{-i}) Q^i(x_{t+1}^i|x_t^i, a_t)$$
$$Q^{-i}(x_{t+1}^{-i}|x_t^{-i}, a_t) P^{\beta_{t+1:T}^i \beta_{t+1:T}^{*,-i},\, \mu_{t+1}^*[a_{1:t}]}(a_{t+1:T}, x_{t+2:T} \big| a_{1:t}, x_{1:t}^i, x_{t+1}) \tag{54c}$$

where (54c) follows from the conditional independence of types given common information, as shown in Claim 1, and the fact that probability on $(a_{t+1:T}, x_{2+t:T})$ given $a_{1:t}, x^i_{1:t-1}, x_{t:t+1}, \mu^*_t[a_{1:t-1}]$ depends on $a_{1:t}, x^i_{1:t}, x_{t+1}, \mu^*_{t+1}[a_{1:t}]$ through $\beta^i_{t+1:T}\beta^{*,-i}_{t+1:T}$. Similarly, the denominator in (54a) is given by

$$
\begin{aligned}
Dr =& \sum_{\tilde{x}^{-i}_t} P^{\beta^i_{t:T}\beta^{*,-i}_{t:T},\mu^*_t}(\tilde{x}^{-i}_t|a_{1:t-1}, x^i_{1:t})\beta^i_t(a^i_t|a_{1:t-1}, x^i_{1:t})\beta^{*,-i}_t(a^{-i}_t|a_{1:t-1}, \tilde{x}^{-i}_t)Q^i(x^i_{t+1}|x^i_t, a_t) && \text{(54d)}\\
=& \sum_{\tilde{x}^{-i}_t} \mu^{*,-i}_t[a_{1:t-1}](\tilde{x}^{-i}_t)\beta^i_t(a^i_t|a_{1:t-1}, x^i_{1:t})\beta^{*,-i}_t(a^{-i}_t|a_{1:t-1}, \tilde{x}^{-i}_t)Q^i(x^i_{t+1}|x^i_t, a_t) && \text{(54e)}
\end{aligned}
$$

By canceling the terms $\beta^i_t(\cdot)$ and $Q^i(\cdot)$ in the numerator and the denominator, (54a) is given by

$$
\begin{aligned}
&\frac{\sum_{x^{-i}_t} \mu^{*,-i}_t[a_{1:t-1}](x^{-i}_t)\beta^{*,-i}_t(a^{-i}_t|a_{1:t-1}, x^{-i}_t)Q^{-i}_{t+1}(x^{-i}_{t+1}|x^{-i}_t, a_t)}{\sum_{\tilde{x}^{-i}_t} \mu^{*,-i}_t[a_{1:t-1}](\tilde{x}^{-i}_t)\beta^{*,-i}_t(a^{-i}_t|a_{1:t-1}, \tilde{x}^{-i}_t)}\times\\
&P^{\beta^i_{t+1:T}\beta^{*,-i}_{t+1:T},\mu^*_{t+1}[a_{1:t}]}(a_{t+1:T}, x_{t+2:T}|a_{1:t}, x^i_{1:t}, x_{t+1}) && \text{(54f)}\\
=&\mu^{*,-i}_{t+1}[a_{1:t}](x^{-i}_{t+1})P^{\beta^i_{t+1:T}\beta^{*,-i}_{t+1:T},\mu^*_{t+1}[a_{1:t}]}(a_{t+1:T}, x_{t+2:T}|a_{1:t}, x^i_{1:t}, x_{t+1}) && \text{(54g)}\\
=&P^{\beta^i_{t+1:T}\beta^{*,-i}_{t+1:T},\mu^*_{t+1}[a_{1:t}]}(x^{-i}_{t+1}, a_{t+1:T}, x_{t+2:T}|a_{1:t}, x^i_{1:t+1}), && \text{(54h)}
\end{aligned}
$$

where (54g) follows from using the definition of $\mu^{*,-i}_{t+1}[a_{1:t}](x^{-i}_t)$ in the forward recursive step in (14) and the definition of the belief update in (37).

■

Lemma 3: $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t-1}, x^i_{1:t}) \in \mathcal{H}^i_t$,

$$
V^i_t(\underline{\mu}^*_t[a_{1:t-1}], x^i_t) = \mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T},\mu^*_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^i(X_n, A_n)\big|a_{1:t-1}, x^i_{1:t}\right\}. \tag{55}
$$

Proof:

We prove the Lemma by induction. For $t = T$,

$$
\begin{aligned}
&\mathbb{E}^{\beta^{*,i}_T\beta^{*,-i}_T,\mu^*_T[a_{1:T-1}]}\left\{R^i(X_T, A_T)\big|a_{1:T-1}, x^i_{1:T}\right\}\\
&= \sum_{x^{-i}_T a_T} R^i(x_T, a_T)\mu^*_T[a_{1:T-1}](x^{-i}_T)\beta^{*,i}_T(a^i_T|a_{1:T-1}, x^i_T)\beta^{*,-i}_T(a^{-i}_T|a_{1:T-1}, x^{-i}_T) && \text{(56a)}\\
&= V^i_T(\underline{\mu}^*_T[a_{1:T-1}], x^i_T), && \text{(56b)}
\end{aligned}
$$

where (56b) follows from the definition of V^i_t in (9) and the definition of β^*_T in the forward recursion in (13).

Suppose the claim is true for $t+1$, i.e., $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t}, x^i_{1:t+1}) \in \mathcal{H}^i_{t+1}$

$$
V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t}], x^i_{t+1}) = \mathbb{E}^{\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T},\mu^*_{t+1}[a_{1:t}]}\left\{\sum_{n=t+1}^{T} R^i(X_n, A_n)\big|a_{1:t}, x^i_{1:t+1}\right\}. \tag{57}
$$

Then $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t-1}, x^i_{1:t}) \in \mathcal{H}^i_t$, we have

$$
\begin{aligned}
&\mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R^i(X_n, A_n) \big| a_{1:t-1}, x^i_{1:t} \right\} \\
&= \mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \big\{ R^i(X_t, A_t) + \\
&\qquad \mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \big| a_{1:t-1}, A_t, x^i_{1:t}, X^i_{t+1} \right\} \big| a_{1:t-1}, x^i_{1:t} \right\} && \text{(58a)} \\
&= \mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \big\{ R^i(X_t, A_t) + \\
&\qquad \mathbb{E}^{\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T}, \mu^*_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \big| a_{1:t-1}, A_t, x^i_{1:t}, X^i_{t+1} \right\} \big| a_{1:t-1}, x^i_{1:t} \right\} && \text{(58b)} \\
&= \mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \left\{ R^i(X_t, A_t) + V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t-1}A_t], X^i_{t+1}) \big| a_{1:t-1}, x^i_{1:t} \right\} && \text{(58c)} \\
&= \mathbb{E}^{\beta^{*,i}_{t}\beta^{*,-i}_{t}, \mu^*_t[a_{1:t-1}]} \left\{ R^i(X_t, A_t) + V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t-1}A_t], X^i_{t+1}) \big| a_{1:t-1}, x^i_{1:t} \right\} && \text{(58d)} \\
&= V^i_t(\underline{\mu}^*_t[a_{1:t-1}], x^i_t), && \text{(58e)}
\end{aligned}
$$

where (58b) follows from Lemma 2 in Appendix D, (58c) follows from the induction hypothesis in (57), (58d) follows because the random variables involved in expectation, X^{-i}_t, A_t, X^i_{t+1} do not depend on $\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T}$ and (58e) follows from the definition of β^*_t in the forward recursion in (13), the definition of μ^*_{t+1} in (14) and the definition of V^i_t in (9). ■

References

[1] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, ser. MIT Press Books. The MIT Press, 1994, vol. 1.

[2] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. Academic Press, 1982.

[3] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[4] E. Maskin and J. Tirole, “Markov perfect equilibrium: I. observable actions,” *Journal of Economic Theory*, vol. 100, no. 2, pp. 191–219, 2001.

[5] A. Nayyar, A. Gupta, C. Langbort, and T. Başar, “Common information based markov perfect equilibria for stochastic games with asymmetric information: Finite games,” *IEEE Trans. Automatic Control*, vol. 59, no. 3, pp. 555–570, March 2014.

[6] A. Gupta, A. Nayyar, C. Langbort, and T. Basar, “Common information based markov perfect equilibria for linear-gaussian games with asymmetric information,” *SIAM Journal on Control and Optimization*, vol. 52, no. 5, pp. 3228–3260, 2014.

[7] Y. Ouyang, H. Tavafoghi, and D. Teneketzis, “Dynamic oligopoly games with private Markovian dynamics,” in *Proc. 54th IEEE Conf. Decision and Control (CDC)*, 2015.

[8] Y.-C. Ho, “Team decision theory and information structures,” *Proceedings of the IEEE*, vol. 68, no. 6, pp. 644–654, 1980.

[9] A. Nayyar, A. Mahajan, and D. Teneketzis, “Decentralized stochastic control with partial history sharing: A common information approach,” *Automatic Control, IEEE Transactions on*, vol. 58, no. 7, pp. 1644–1658, 2013.

[10] A. Mahajan, “Optimal decentralized control of coupled subsystems with control sharing,” *Automatic Control, IEEE Transactions on*, vol. 58, no. 9, pp. 2377–2382, 2013.

[11] A. Mahajan and D. Teneketzis, “On the design of globally optimal communication strategies for real-time communcation systems with noisy feedback,” *IEEE J. Select. Areas Commun.*, no. 4, pp. 580–595, May 2008.

[12] A. Nayyar and D. Teneketzis, “On globally optimal real-time encoding and decoding strategies in multi-terminal communication systems,” in *Proc. IEEE Conf. on Decision and Control*, Cancun, Mexico, Dec. 2008, pp. 1620–1627.

[13] D. Vasal and A. Anastasopoulos, “Stochastic control of relay channels with cooperative and strategic users,” *Communications, IEEE Transactions on*, vol. 62, no. 10, pp. 3434–3446, Oct 2014.